Inverse material search and synthesis verification by hand drawings *via* transfer learning and contour detection

Nikita Serov[1] and Vladimir Vinogradov[1]*

[1] International Institute "Solution Chemistry of Advanced Materials and Technologies", ITMO University, Saint-Petersburg, 191002, Russian Federation

Abstract: Nanomaterials of various morphologies and chemistry have an extensive use as photonic devices, advanced catalysts, sorbents for water purification, agrochemicals, platforms for drug delivery as well as imaging systems to name a few. However, search for synthesis routes giving custom nanomaterials for particular needs with the desired structure, shape, and size remains a challenge and is often implemented by manual research articles screening. Here, we develop for the first time scanning and transmission electron microscopy (SEM/TEM) reverse image search and hand drawing-based search *via* transfer learning (TL), namely, VGG16 convolutional neural network (CNN) repurposing for image features extraction and subsequent image similarity determination. Moreover, we demonstrate case use of this platform on calcium carbonate system, where sufficient amount of data was acquired by random high throughput multiparametric synthesis, as well as on Au nanoparticles (NPs) data extracted from the articles. This approach can be not only used for advanced nanomaterials search and synthesis procedure verification, but also can be further combined with machine learning (ML) solutions to provide data-driven novel nanomaterials discovery.

Nanomaterials are widely used as photonic devices, advanced catalysts, sorbents for water purification, agrochemicals, platforms for drug delivery etc. due to its ability to control the shape, size, morphology, surface chemistry, and composition, which strongly influence its physicochemical properties as well as its biological behavior. Calcium carbonate represents an inorganic material with a huge potential in the formation of complex micro- and macrostructures, [1,2] which is evident from its wide use by the living organisms in the process of biomineralization, [3–5] thereby it is widely used in drug delivery, [6–9] photonics [10,11] etc. At the same time, gold nanomaterials are widespread in drug delivery and photonics mainly due to surface plasmon resonance [12], photothermal activity, [13–15] and surface chemistry tunability, [16] which, coupled with the ability of precise shape control, [17,18] makes it a promising nanomaterial for nanomedicine and physics.

The ever-growing amount of experimental data devoted to nanomaterials properties and its synthesis procedures creates a need for fully systematized data collection, storage, and precise search. Several annotated materials synthesis-related databases exist, [19–21] where the content is usually processed via natural language processing (NLP)-based text mining [22] allowing for either direct or inverse materials search from synthesis procedures to the outcome and *vice versa*, respectively. In the field of materials science, there is a need in inverse materials search since scientists are often puzzled over how to synthesize a material with desired properties prior to what one would get given the set of experimental conditions. To date, there are several solutions toward nanomaterials synthesis search such as, for example, Nano (https://nano.nature.com/) based on machine learning-driven automated procedures extraction from research articles, although their search is limited to the keywords.

Electron microscopy (EM) remains one of the most demanded instruments for materials characterization giving the information about material morphology, size, as well as the shape. Controlling these parameters is of great importance [23] to obtain drug delivery systems (DDSs) with desired biodistribution in the organism, which were shown to depend strongly on DDS size and shape, [24]

photonic crystals with low polydispersity, drastically affecting its optical properties [25,26] etc. Moreover, TEM images give insights into the material crystallinity as well as an internal structuring allowing to study core-shell and hollow structures composed of different crystalline phases. Therefore, EM images represent meaningful synthesis outcomes, which can be used for reverse material search.

The main question is how to distinguish the difference between two or more synthesis outcomes represented by SEM images. Pixel-by-pixel comparison via image distance calculation [27,28] is able to find exactly the same images but fails on other images of the same objects since it does not consider the relationships between the pixels, not to mention its high computational cost. Instead of pixels, image features invariant to some geometric transformations can be used. [29] For instance, Scale Invariant Feature Transform (SIFT) can extract rotation-insensitive located image features on various scales, which then can be used for image similarity calculation *via* nearest neighbors. However, SIFT and other similar algorithms [30,31] suffer from low computation speed and are sensitive to brightness/contrast as well as blurring. CNNs usually outperform such algorithms in feature extraction and subsequent classification tasks while being more robust. [32] In their work, Modarres et al. have implemented TL approach on Inception-v3 model pre-trained on ImageNet 2012 dataset for SEM images supervised classification. [33] Therefore, CNNs can be used for feature extraction and subsequent image similarity determination for reverse image search. Due to the lack of large amounts of experimental data in materials science, it is rather difficult to achieve sufficient training accuracy on only materials science datasets, that is why TL approach, which refers to the use of pre-trained ML models for another task, gains the momentum. In this Article, we develop EM reverse image search based on VGG16 CNN repurposing for automated image features extraction and subsequent image similarity determination. Furthermore, we demonstrate case use of this approach on calcium carbonate system, where sufficient amount of data was acquired *via* random high throughput multiparametric solution chemistry synthesis. Presented approach can be not only used for custom nanomaterials search and synthesis procedure verification, but also can be further equipped with ML solutions to provide data-driven novel nanomaterials discovery.

Results and discussion

To generate meaningful experimental data to form a database of synthesis routes and its outcomes, namely, scanning electron microscope (SEM) images showing micro-/nanoparticle morphology, size, and shape, random high throughput screening, namely was introduced on inorganic calcium carbonate system as a case use including materials synthesis, evaluation using SEM, and database expansion (**Fig. 1a**, **1b**, and **1c**, respectively). In particular, randomization of reagents volumes with fixed stock concentrations, coupled with the association of samples in small arbitrary groups of random synthesis parameters e.g., temperature, synthesis time etc. was implemented. To cover the vast majority of possible materials shapes, sizes, polydispersity, and surface morphologies, such parameters as synthesis time, temperature, stirring rate, concentrations of precursors e.g. calcium, carbonate, and bicarbonate ions, mass fraction of miscible/immiscible solvents e.g. methanol, hexanol, isopropyl alcohol (IPA), dimethylformamide (DMFA), propylene glycol, ethylene glycol (EG), tert-butyl alcohol, charged and uncharged polymers of various molar weights e.g. polyethylene glycol (PEG), polystyrene sulfonate (PSS), polyvinylpyrrolidone (PVP), polyethyleneimine (PEI), polyacrylic acid (PAA), and concentrations of differently-charged surfactants e.g. sodium dodecyl sulfate, Triton X-15, myristyltrimethylammonium bromide, cetrimonium bromide (CTAB), were varied in a wide range as it can be seen from distribution box plots for each of the variables on **Fig. S1**. Randomization of these variables allows to exclude human bias as well as to include 'negative' outcomes, which are very important for any subsequent ML but still under-represented in the majority

of research articles. Accordingly, the database of >200 individual nanomaterials was collected (**Fig. S2**) consisting >20 unique shapes (**Fig. 1d**), and SEM image was assigned to every single synthetic procedure in the database as an outcome.

To achieve reverse image search on SEM images and subsequent image label-based synthesis procedure retrieval from the database, image features extraction needs to be implemented, which is usually achieved by the utilization of encoder-decoder CNNs gradually compressing the image dimensions and trying to reconstruct it with unique features extracted from the images. TL approach has been implemented in this work, namely, re-purposing of the widely used VGG16 CNN model (**Fig. 2a**) pre-trained on >14.000.000 images of macroscopic objects of as many as 20.000 categories for SEM image features extraction (**Fig. 2b**). VGG16 CNN consists of convolution, as well as pooling and fully connected dense layers. All these layers represent the mathematical transformation of the (1,224,224,3)-shaped input image pixel intensities, where convolution basically represents the application of filters to the groups of pixels thereby considering interrelations between the adjacent pixels, pooling compresses the image resulting in compressed image representation, and dense layers are usually used for further classification tasks. The last fully connected layer of shape (,4096) carries 4096 features generated for every single image, which are then compressed to 200 and used for cosine distance (**eq. 1**, cosine similarity of two n-dimensional vectors A and B) determination between the images represented as vectors in 200-dimensional feature space. To demonstrate that this model captures complex crystal morphologies on SEM images, several queries were made resulting in top 3 the most similar images in the database (**Fig. 2c**). For instance, the model was able to find as simple shapes e.g., cubes, spheres, and spikes, as more complex e.g. urchins, flowers, and sphere aggregates even of comparable sizes, since the exact match is limited to the database size. Moreover, it can be seen that shape mixes are also recognized. To check the ability of this algorithm to reflect the shape abundancies as well as particle sizes, these parameters were calculated for query containing both spheres and cubes and cubes only as well as top 3 the closest SEM images in the database (**Fig. 3a** and **3b**, respectively). The algorithm, indeed, reflects sphere abundancy in the query image equal to 71% trying to find the best match, where top 3 similar images have this parameter equal to 87, 48, and 98% (**Fig. 3c**). Moreover, according to the sphere size distribution in query image of 1.66±0.19 µm as well as in top 3 most similar images, 2.15±0.26, 1.88±0.40, and 2.55±0.39 µm (**Fig. 3d**), respectively, query results represent the compromise between particles shape and size. To minimize the impact of shape diversity, image query with cubes only has been examined, where the results of comparable sizes were suggested by the algorithm (**Fig. 3e**), namely, query results containing cubes with side lengths of 5.6 and 4.5, 5.2 and 3.3, 8.35 and 5.2 µm for a query image with these parameters equal to 5.9 and 3.3 µm were obtained. It is important to note that the increased number of crystal defects going from the query image to the 3rd query result is observed (**Fig. 3b**, **inset**), which suggest this algorithm is sensitive to the surface morphology of the material. From the abovementioned, it can be concluded that this approach allows to search for the materials with the closest shape abundancies, size distributions, as well as material surface morphologies, where all the query results are ranked given the compromise between all of these parameters.

To reveal further material insights captured by the algorithm, image data analysis was implemented following image augmentation through the generation of flipped copies of SEM images existing in the database. Principal component analysis (PCA) approach performed on the features extracted

$$Eq.\,1 \quad \text{cosine similarity} = \cos(\theta) = \frac{A \times B}{||A|| \times ||B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}} = 1 - cosine\ distance$$

from the SEM images has shown the ability of the most representative shapes e.g. cubes, spheres, and spikes, to form distinguishable clusters (**Fig. S3a**). PCA is a dimension reduction technique, thus some valuable information in the form of feature variance may be lost when n-dimensional space is compressed to visualizable 2-dimensional one. Cumulative explained variance ratio (CEVR) of 2 principal components equal to 28% suggests that the majority of data variance is preserved during the transformation (**Fig. S4**). K-nearest neighbors (kNN) algorithm implemented on full augmented SEM image dataset allowed to also identify 3 distinctive clusters, which does not correspond with the number of shapes presented in the dataset. These findings, together with the existence of calcium carbonate in 3 main stable crystalline phases as well as literature data indicating spheres are usually consist of vaterite phase, [1,34] cubes – of calcite, [35,36] and spikes – of aragonite, [35,37] suggest that this algorithm probably not only detect the material shape, size, and surface morphology, but also the crystalline phase, where the latter dictates the listed parameters. Therefore, this 2-dimensional scatter plot can be potentially interpreted as a material phase diagram, however, more thorough investigations are needed, which is out of the article scope.

To show the versatility of the developed approach as well as its indifference to the material used, its size, as well as the type of the image, this concept was verified on transmission electron microscopy (TEM) images of gold nanoparticles (Au NPs) of six shapes e.g., sphere, cube, rod, dumbbell, trigonal, and amorphous (**Fig. 4a**) widely used mainly due to their shape-dependent surface plasmon resonance and optical properties, manually extracted from the articles. Algorithm was able to find the most similar images in the collected set of 15 TEM images (**Fig. 4b**). For instance, all rod-like shapes presented in the image set were found, while the 3rd query result is turned out to be the best of the worst having the value of cosine distance equal to 1.00, while this parameter of the 1st and 2nd query results is equal to 0.37 and 0.52, respectively. The big cosine distance between the query image and 1st result can be explained by the big difference in NPs lengths (74.3±6.2 and 63.7±5.5 nm, respectively) and widths (19.8±1.4 and 11.9±1.2 nm, respectively). Moreover, size sorting of 2 query results containing trigonal and spherical Au NPs of different mean size 14.7±2.1 and 9.5±1.4 nm was observed (**Fig. 4c**), where scale bars were included in the images, thereby being included in the image features, and considered during image similarity determination.

To make step beyond the synthesis verification towards the customized inverse material queries, drawing-based inverse material search was demonstrated for the first time. First, Canny contour detection was implemented on the set of pre-processed with contrasting and Gaussian blurring calcium carbonate-based nanomaterials SEM images to generate hand drawing-like images for further image similarity determination (**Fig. 5a**). To examine, whether this approach is feasible, two queries comprising simple crystal shapes, namely, spheres and cubes, were made (**Fig. 5b**). It is important to note that the algorithm was able to find the closest SEM image in the dataset for a given query with spheres. Moreover, it can be seen that going from the 1st to the very last query result is accompanied by the change in sphere morphology as well as its size and shape (**Fig. 5b**, **inset**) becoming less and less similar to the hand drawn query. More complex query comprising cubes with surface defects has also resulted in a successful search for similar samples even of close sizes, where facet defects were changing from the 1st to the last query result. Therefore, hand drawing-based inverse material search is demonstrated.

Hence, in this study, a novel approach towards the synthesis verification by inverse EM image search and customized drawing-based material query for custom inverse material search is introduced for the first time. TL, namely, VGG16 CNN pre-trained on >14 million images re-purposing was implemented for SEM/TEM image feature extraction and subsequent image similarity determination. Case use of this approach on >200 manually synthesized by random high-throughput

screening calcium carbonate-based nanomaterials of >20 various shapes, sizes, and surface morphologies, as well as on Au NPs of >6 shapes extracted from the research articles was demonstrated, thereby proving approach versatility. It was shown that Canny contour detection enables one to implement hand drawing-based query introducing customized inverse material search with the desired shapes, sizes, and surface morphologies. Developed approach can be not only utilized for advanced nanomaterials search and synthesis procedure verification, but also can be further equipped with machine learning (ML) solutions to provide data-driven novel nanomaterials discovery.

ABBREVIATIONS
CEVR, cumulative explained variance ratio; CNN, convolutional neural network; CTAB, cetrimonium bromide; DDS, drug delivery system; DMFA, dimethylformamide; EG, ethylene glycol; EM, electron microscopy; IPA, isopropyl alcohol; kNN, k-nearest neighbors; ML, machine learning; NLP, natural language processing; NP, nanoparticle; PAA, polyacrylic acid; PCA, principal component analysis; PEG, polyethylene glycol; PEI, polyethylene imine; PSS, polystyrene sulfonate; PVP, polyvinyl pyrrolidone; SEM, scanning electron microscope; SIFT, scale invariant feature transform; TEM, transmission electron microscopy; TL, transfer learning.

DATA AVAILABILITY
Database of synthetic procedures, corresponding SEM images, as well as all the Python code are available on GitHub repository ().

AUTHOR INFORMATION

Corresponding Authors
* Vladimir V. Vinogradov: vinogradov@scamt-itmo.ru

Present Addresses
[1] International Institute "Solution Chemistry of Advanced Materials and Technologies", ITMO University, Saint-Petersburg, 191002, Russian Federation

Funding Sources

REFERENCES

(1)   Guo, X.-H.; Yu, S.-H.; Cai, G.-B. Crystallization in a Mixture of Solvents by Using a Crystal Modifier: Morphology Control in the Synthesis of Highly Monodisperse CaCO3 Microspheres. *Angew. Chemie Int. Ed.* **2006**, *45* (24), 3977–3981. https://doi.org/10.1002/anie.200600029.

(2)   Jiang, W.; Pacella, M. S.; Athanasiadou, D.; Nelea, V.; Vali, H.; Hazen, R. M.; Gray, J. J.; McKee, M. D. Chiral Acidic Amino Acids Induce Chiral Hierarchical Structure in Calcium Carbonate. *Nat. Commun.* **2017**, *8* (1), 15066. https://doi.org/10.1038/ncomms15066.

(3)   Dhami, N. K.; Reddy, M. S.; Mukherjee, A. Biomineralization of Calcium Carbonates and Their Engineered Applications: A Review. *Front. Microbiol.* **2013**, *4*. https://doi.org/10.3389/fmicb.2013.00314.

(4)   Cartwright, J. H. E.; Checa, A. G.; Gale, J. D.; Gebauer, D.; Sainz-Díaz, C. I. Calcium Carbonate Polyamorphism and Its Role in Biomineralization: How Many Amorphous Calcium Carbonates Are There? *Angew. Chemie Int. Ed.* **2012**, *51* (48), 11960–11970. https://doi.org/10.1002/anie.201203125.

(5)   Yao, S.; Jin, B.; Liu, Z.; Shao, C.; Zhao, R.; Wang, X.; Tang, R. Biomineralization: From Material Tactics to Biological Strategy. *Adv. Mater.* **2017**, *29* (14), 1605903. https://doi.org/10.1002/adma.201605903.

(6)   Maleki Dizaj, S.; Barzegar-Jalali, M.; Zarrintan, M. H.; Adibkia, K.; Lotfipour, F. Calcium Carbonate Nanoparticles as Cancer Drug Delivery System. *Expert Opin. Drug Deliv.* **2015**, *12* (10), 1649–1660. https://doi.org/10.1517/17425247.2015.1049530.

(7)   Serov, N.; Prilepskii, A.; Sokolov, A.; Vinogradov, V. Synthesis of Plasmin-Loaded Fe 3 O 4 @CaCO 3 Nanoparticles: Towards Next-Generation Thrombolytic Drugs. *ChemNanoMat* **2019**, *5* (10), 1267–1271. https://doi.org/10.1002/cnma.201900359.

(8)   Zhao, Y.; Luo, Z.; Li, M.; Qu, Q.; Ma, X.; Yu, S.-H.; Zhao, Y. A Preloaded Amorphous Calcium Carbonate/Doxorubicin@Silica Nanoreactor for PH-Responsive Delivery of an Anticancer Drug. *Angew. Chemie Int. Ed.* **2015**, *54* (3), 919–922. https://doi.org/10.1002/anie.201408510.

(9)   Serov, N.; Darmoroz, D.; Lokteva, A.; Chernyshov, I.; Koshel, E.; Vinogradov, V. One-Pot Synthesis of Template-Free Hollow Anisotropic CaCO 3 Structures: Towards Inorganic Shape-Mimicking Drug Delivery Systems. *Chem. Commun.* **2020**, *56* (80), 11969–11972. https://doi.org/10.1039/D0CC05502F.

(10)   Lee, K.; Wagermaier, W.; Masic, A.; Kommareddy, K. P.; Bennet, M.; Manjubala, I.; Lee, S.-W.; Park, S. B.; Cölfen, H.; Fratzl, P. Self-Assembly of Amorphous Calcium Carbonate Microlens Arrays. *Nat. Commun.* **2012**, *3* (1), 725. https://doi.org/10.1038/ncomms1720.

(11)   Li, C.; Hong, G.; Yu, H.; Qi, L. Facile Fabrication of Honeycomb-Patterned Thin Films of Amorphous Calcium Carbonate and Mosaic Calcite. *Chem. Mater.* **2010**, *22* (10), 3206–3211. https://doi.org/10.1021/cm100363a.

(12)   Amendola, V.; Pilot, R.; Frasconi, M.; Maragò, O. M.; Iatì, M. A. Surface Plasmon Resonance in Gold Nanoparticles: A Review. *J. Phys. Condens. Matter* **2017**, *29* (20), 203002. https://doi.org/10.1088/1361-648X/aa60f3.

(13)   Jiang, K.; Smith, D. A.; Pinchuk, A. Size-Dependent Photothermal Conversion Efficiencies of Plasmonically Heated Gold Nanoparticles. *J. Phys. Chem. C* **2013**, *117* (51), 27073–27080. https://doi.org/10.1021/jp409067h.

(14)   Wang, S.; Chen, K.-J.; Wu, T.-H.; Wang, H.; Lin, W.-Y.; Ohashi, M.; Chiou, P.-Y.; Tseng, H.-R. Photothermal Effects of Supramolecularly Assembled Gold Nanoparticles for the Targeted Treatment of Cancer Cells. *Angew. Chemie* **2010**, *122* (22), 3865–3869. https://doi.org/10.1002/ange.201000062.

(15)   Vines, J. B.; Yoon, J.-H.; Ryu, N.-E.; Lim, D.-J.; Park, H. Gold Nanoparticles for Photothermal Cancer Therapy. *Front. Chem.* **2019**, *7*. https://doi.org/10.3389/fchem.2019.00167.

(16)   Zeng, S.; Yong, K.-T.; Roy, I.; Dinh, X.-Q.; Yu, X.; Luan, F. A Review on Functionalized Gold Nanoparticles for Biosensing Applications. *Plasmonics* **2011**, *6* (3), 491–506. https://doi.org/10.1007/s11468-011-9228-1.

(17)   Personick, M. L.; Langille, M. R.; Zhang, J.; Mirkin, C. A. Shape Control of Gold Nanoparticles by Silver Underpotential Deposition. *Nano Lett.* **2011**, *11* (8), 3394–3398. https://doi.org/10.1021/nl201796s.

(18)   Lohse, S. E.; Murphy, C. J. The Quest for Shape Control: A History of Gold Nanorod Synthesis. *Chem. Mater.* **2013**, *25* (8), 1250–1261. https://doi.org/10.1021/cm303708p.

(19)   Kim, E.; Huang, K.; Tomala, A.; Matthews, S.; Strubell, E.; Saunders, A.; McCallum, A.; Olivetti, E. Machine-Learned and Codified Synthesis Parameters of Oxide Materials. *Sci. Data* **2017**, *4* (1), 170127. https://doi.org/10.1038/sdata.2017.127.

(20)   Kononova, O.; Huo, H.; He, T.; Rong, Z.; Botari, T.; Sun, W.; Tshitoyan, V.; Ceder, G. Text-Mined Dataset of Inorganic Materials Synthesis Recipes. *Sci. Data* **2019**, *6* (1). https://doi.org/10.1038/s41597-019-0224-1.

(21)   Kim, E.; Huang, K.; Saunders, A.; McCallum, A.; Ceder, G.; Olivetti, E. Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning. *Chem. Mater.* **2017**, *29* (21), 9436–9444. https://doi.org/10.1021/acs.chemmater.7b03500.

(22) Krallinger, M.; Rabal, O.; Lourenço, A.; Oyarzabal, J.; Valencia, A. Information Retrieval and Text Mining Technologies for Chemistry. *Chem. Rev.* **2017**, *117* (12), 7673–7761. https://doi.org/10.1021/acs.chemrev.6b00851.

(23) Baalousha, M.; Lead, J. R. Nanoparticle Dispersity in Toxicology. *Nat. Nanotechnol.* **2013**, *8* (5), 308–309. https://doi.org/10.1038/nnano.2013.78.

(24) Wei, Y.; Quan, L.; Zhou, C.; Zhan, Q. Factors Relating to the Biodistribution & Clearance of Nanoparticles & Their Effects on in Vivo Application. *Nanomedicine* **2018**, *13* (12), 1495–1512. https://doi.org/10.2217/nnm-2018-0040.

(25) Scholes, G. D. Controlling the Optical Properties of Inorganic Nanoparticles. *Adv. Funct. Mater.* **2008**, *18* (8), 1157–1172. https://doi.org/10.1002/adfm.200800151.

(26) Allard, M.; Sargent, E. H. Impact of Polydispersity on Light Propagation in Colloidal Photonic Crystals. *Appl. Phys. Lett.* **2004**, *85* (24), 5887–5889. https://doi.org/10.1063/1.1835533.

(27) Moghaddam, B.; Nastar, C.; Pentland, A. A Bayesian Similarity Measure for Direct Image Matching. In *Proceedings of 13th International Conference on Pattern Recognition*; IEEE, 1996; pp 350–358 vol.2. https://doi.org/10.1109/ICPR.1996.546848.

(28) Di Gesù, V.; Starovoitov, V. Distance-Based Functions for Image Comparison. *Pattern Recognit. Lett.* **1999**, *20* (2), 207–214. https://doi.org/10.1016/S0167-8655(98)00115-9.

(29) Nixon, M. S.; Aguado, A. S. *Feature Extraction and Image Processing for Computer Vision*; 2019. https://doi.org/10.1016/C2017-0-02153-5.

(30) Dellinger, F.; Delon, J.; Gousseau, Y.; Michel, J.; Tupin, F. SAR-SIFT: A SIFT-Like Algorithm for SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53* (1), 453–466. https://doi.org/10.1109/TGRS.2014.2323552.

(31) Khan, N. Y.; McCane, B.; Wyvill, G. SIFT and SURF Performance Evaluation against Various Image Deformations on Benchmark Dataset. In *2011 International Conference on Digital Image Computing: Techniques and Applications*; IEEE, 2011; pp 501–506. https://doi.org/10.1109/DICTA.2011.90.

(32) Zheng, L.; Yang, Y.; Tian, Q. SIFT Meets CNN: A Decade Survey of Instance Retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40* (5), 1224–1244. https://doi.org/10.1109/TPAMI.2017.2709749.

(33) Modarres, M. H.; Aversa, R.; Cozzini, S.; Ciancio, R.; Leto, A.; Brandino, G. P. Neural Network for Nanoscience Scanning Electron Microscope Image Recognition. *Sci. Rep.* **2017**, *7* (1), 13282. https://doi.org/10.1038/s41598-017-13565-z.

(34) Wang, Y.; Moo, Y. X.; Chen, C.; Gunawan, P.; Xu, R. Fast Precipitation of Uniform CaCO3 Nanospheres and Their Transformation to Hollow Hydroxyapatite Nanospheres. *J. Colloid Interface Sci.* **2010**, *352* (2), 393–400. https://doi.org/10.1016/j.jcis.2010.08.060.

(35) Wang, C. Control the Polymorphism and Morphology of Calcium Carbonate Precipitation from a Calcium Acetate and Urea Solution. *Mater. Lett.* **2008**, *62* (16), 2377–2380. https://doi.org/10.1016/j.matlet.2007.12.020.

(36) Ouhenia, S.; Chateigner, D.; Belkhir, M. A.; Guilmeau, E.; Krauss, C. Synthesis of Calcium Carbonate Polymorphs in the Presence of Polyacrylic Acid. *J. Cryst. Growth* **2008**, *310* (11), 2832–2841. https://doi.org/10.1016/j.jcrysgro.2008.02.006.

(37) Zhang, Z.; Xie, Y.; Xu, X.; Pan, H.; Tang, R. Transformation of Amorphous Calcium Carbonate into Aragonite. *J. Cryst. Growth* **2012**, *343* (1), 62–67. https://doi.org/10.1016/j.jcrysgro.2012.01.025.
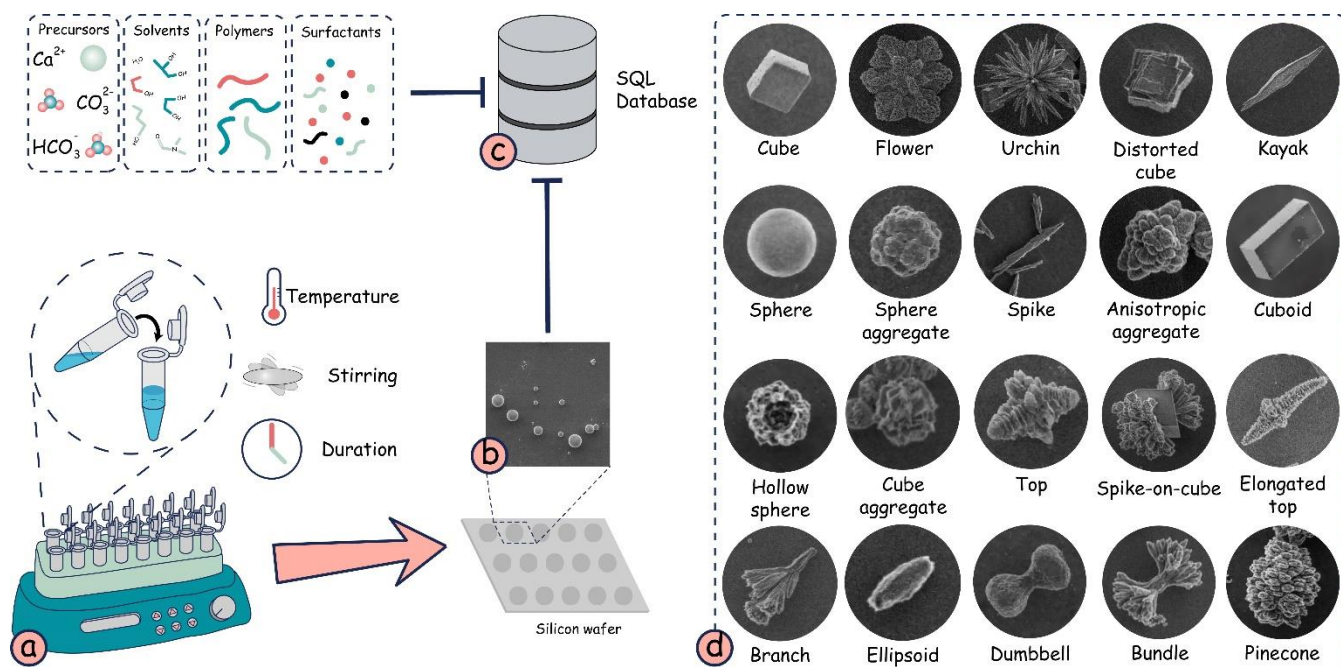
Figure 1. Random high throughput multiparametric synthesis of calcium carbonate-based nanomaterials and database development.
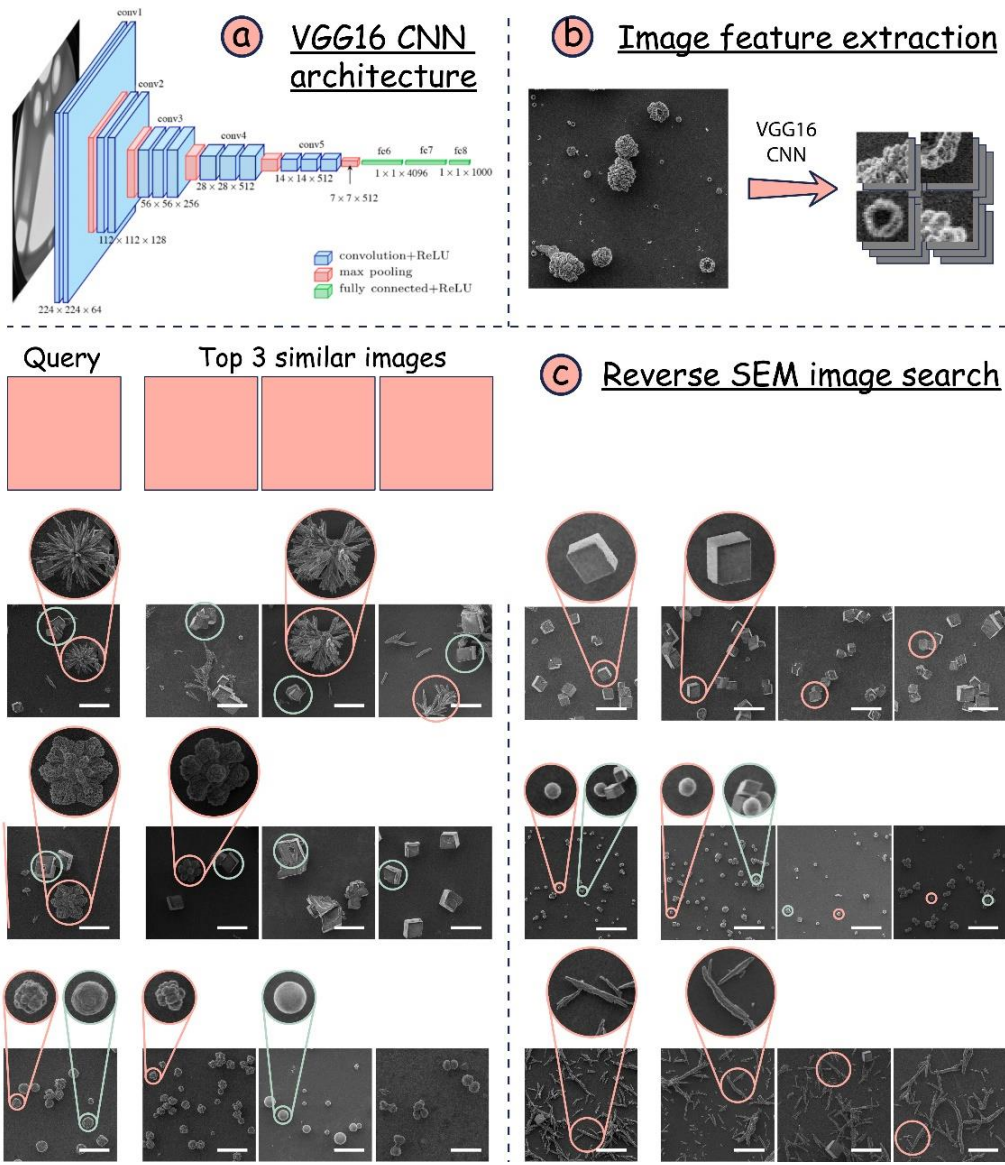
Figure 2. VGG16-based transfer learning for image features extraction and subsequent reverse SEM image search based on image similarity.
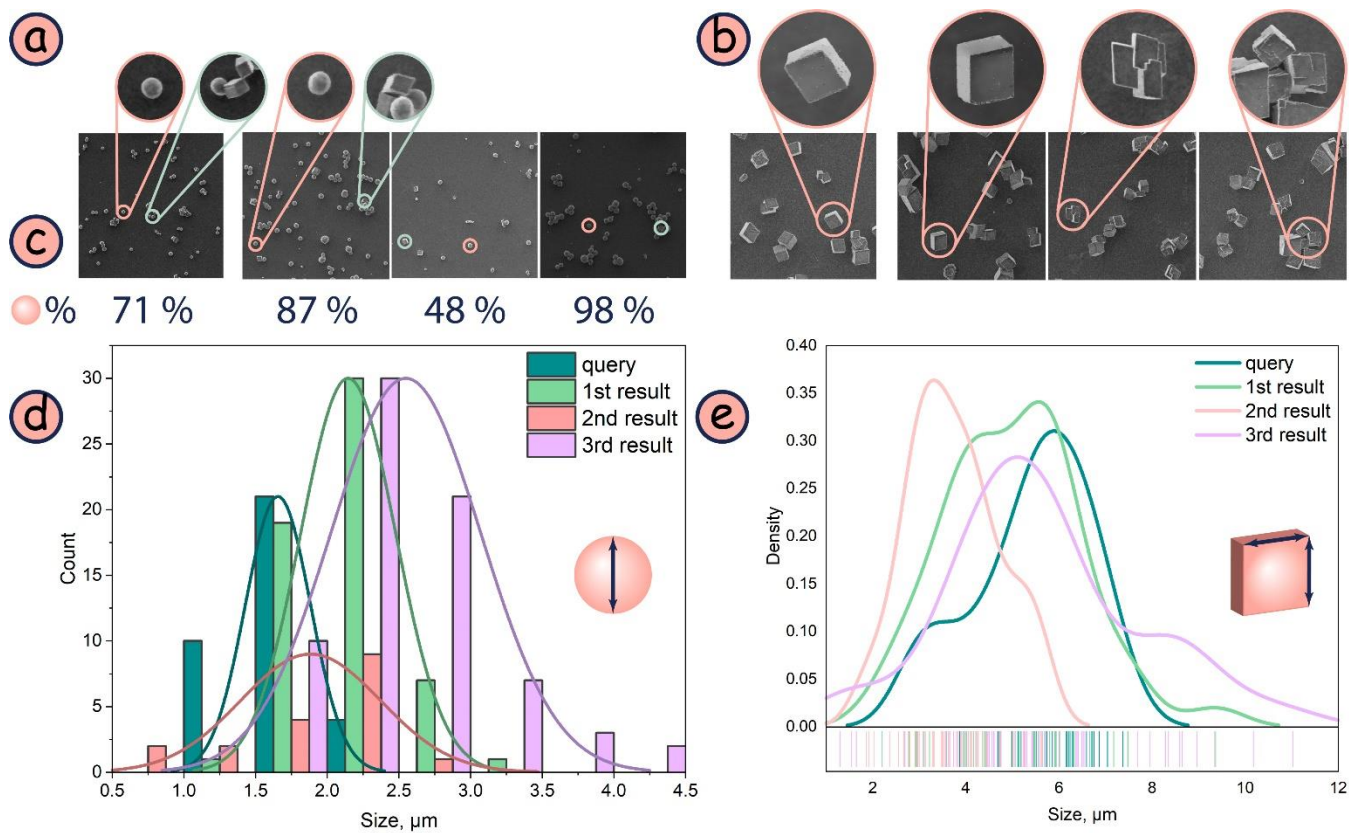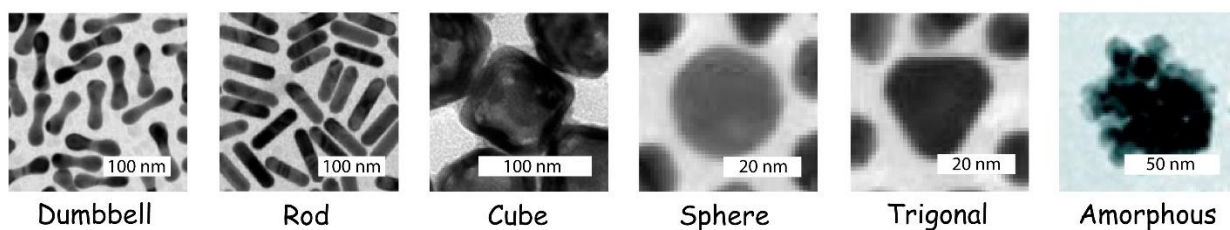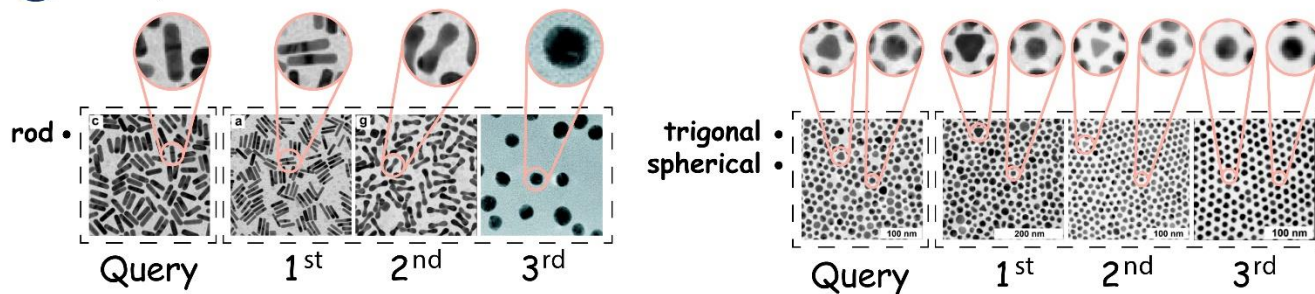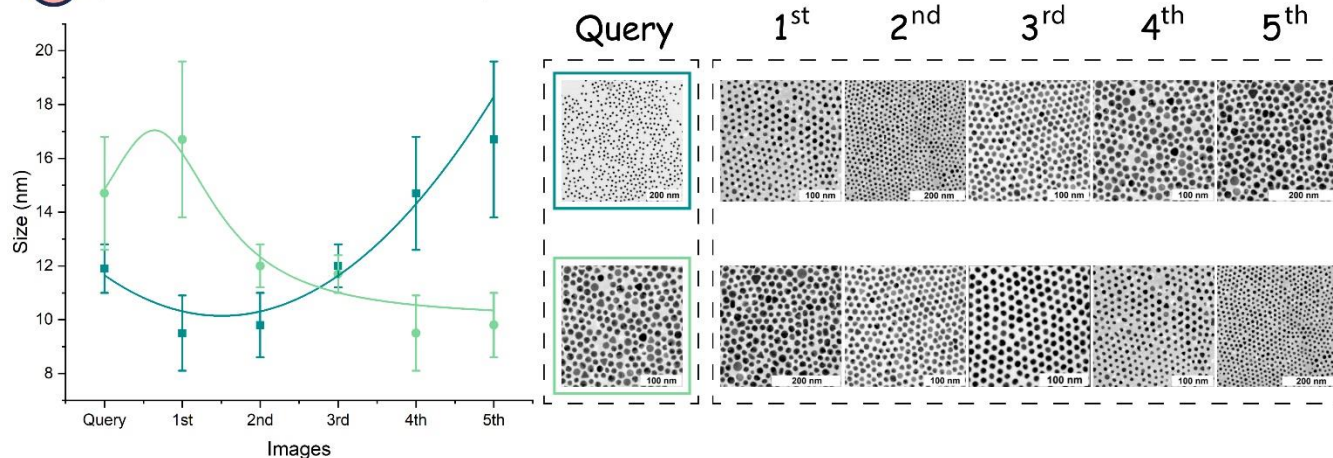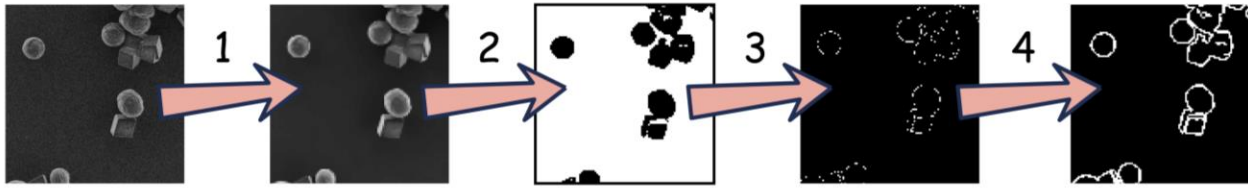
Figure 3. Size distribution of the closest samples based on SEM image similarity.

Figure 4. Proof-of-concept demonstration of algorithm work on differently sized Au NPs of six shapes (a), namely, shape (b) and size (c) differentiation.
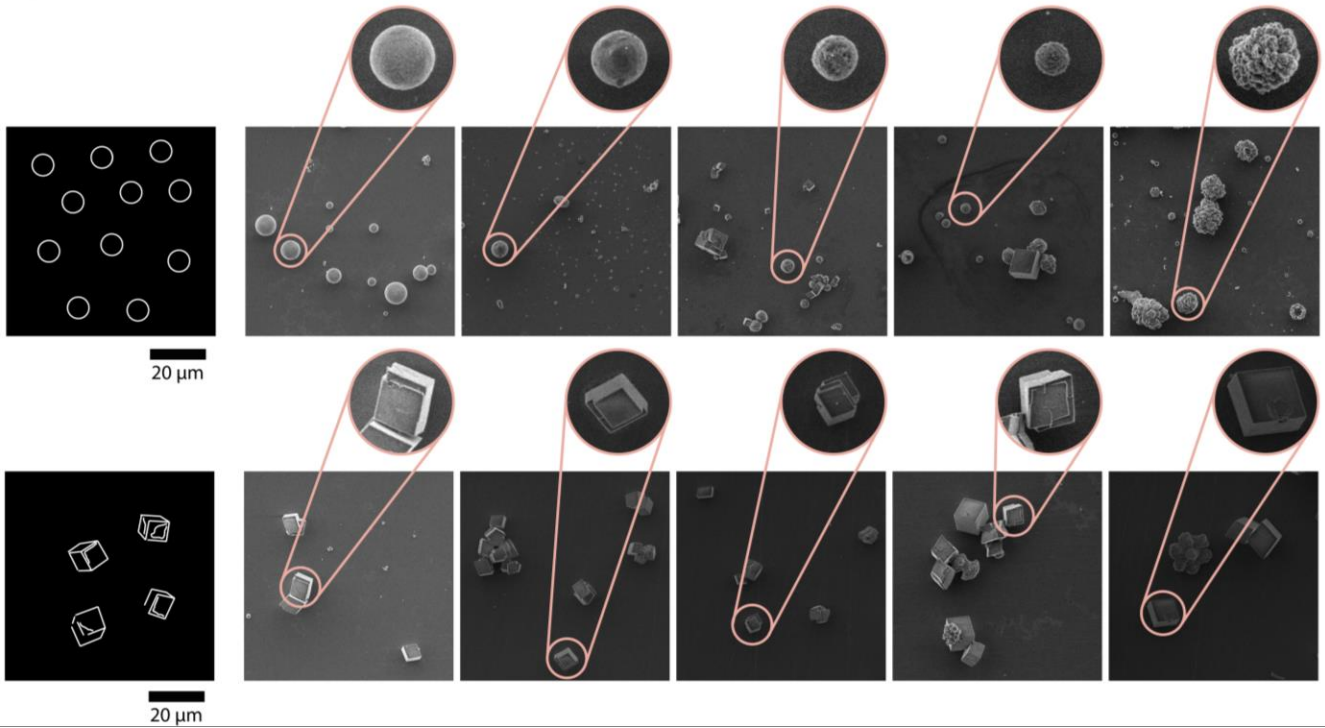
Figure 5. Hand drawing-based reverse image search. A) contour detection implementation (1 – contrasting and denoising; 2 – binarization; 3 – edge detection; 4 – contour detection) and B) reverse search on hand drawn spheres and cubes.